ABSTRACT
            An instrument to measure the computer literacy of
science teachers in grades K-12 is being developed. This instrument,
the Test of Computer Literacy for Science Teachers (TCLST), is part
of ENLIST Micros (a project to develop a curriculum for training
science teachers to use computers). The instrument, based on
essential competencies for computer literacy for science teachers
developed and validated for this project, will be used to evaluate
the effectiveness of the ENLIST Micros curriculum at developing those
essential competencies in science teachers. In addition, TCLST can be
used as a diagnostic test by universities and school districts to
determine if teachers have previously achieved computer literacy and
also for various purposes by researchers. A 12-step procedure for
developing and validating criterion-referenced tests is being used to
develp the TCLST. Results of the first eight steps are reported and
discussed: preliminary considerations; review of objectives; item
writing; assessment of content validity; revisions to test items;
field test administration; revision of test items; and test assembly.
The test is being piloted with over 200 preservice and inservice
science teachers in 10 sites. Results of the pilot study will be used
to establish instrument reliability and its construct and decision
validity. (Author/JN)

The Development of a Test of Computer Literacy for Science Teachers in Grades K-12

James D. Ellis
BSCS
1115 North Cascade Ave.
Colorado Springs, CO 80903

Paul J. Kuerbis
Education Department
The Colorado College
Colorado Springs, CO 80903

The Development of a Test of Computer Literacy for Science Teachers in Grades K-12

Purpose

In this study, the authors are developing a test to measure the computer literacy of science teachers in grades K-12. This instrument is part of ENLIST Micros--a NSF funded project to develop a curriculum for training science teachers to use the computer. The instrument is based on the essential competencies in computer literacy for science teachers developed and validated for this project. The instrument will be used to evaluate the effectiveness of the ENLIST Micros curriculum at developing those essential competencies in science teachers and will be available to researchers who are studying the computer literacy of science teachers.

Theoretical Basis

Science teachers must be computer literate to use the microcomputer as an instructional tool, to introduce students to using the microcomputer to solve problems in science, to facilitate the development of computer literacy by students, to use the microcomputer as a tool to improve the management of instruction, and to develop and exhibit positive attitudes and values toward computer use. Therefore, if we are to succeed in training students for the information age, we must train science teachers to use the computer in instruction, just as they have learned to use textbooks, films, television, and the overhead projector.

However, as Watson (1983) has pointed out, few science teachers have been trained in how to use the computer for instructional purposes or participated in instruction using a computer. Although some teacher education programs may help science teachers learn about the computer, all teacher educators face difficult decisions as they attempt to decide, for example, "whether teachers should be familiar with computer simulations or be able to design simulations. That is, should science teachers know merely where to obtain computer software or should they know how to improve inadequate software?" (Taylor, 1984, p. 270).

What are the essential competencies in computer literacy for teachers? Furthermore, what might constitute a curriculum for training science teachers to use the computer in the classroom?

ENLIST Micros is seeking answers to those questions. The essential competencies were identified as a precursor to the development of the curriculum for training science teachers to use the microcomputer for instruction (Ellis and Kuerbis, 1985). An advisory committee consisting of science teachers, experts in educational computing in the sciences, school administrators, and publishers of science software reviewed the list of essential competencies and recommended the structure and organization for the curriculum. During the summer and fall 1985, the curriculum was prepared by a group of six writers, two computer programmers, three media designers, an artist, a secretary, and the two co-directors.

The curriculum was field tested during the fall of 1985 and the spring of 1986 in more than ten settings (both preservice and

inservice). The experimental materials was thoroughly evaluated
in these settings and will be revised during the summer of 1986.
Then the curriculum will be distributed by a commercial publisher.

As part of the evaluation procedure, a search was conducted
to locate an appropriate measure of computer literacy as defined
by the ENLIST Micros project. No instruments were found that
were closely aligned with the 19 cognitive objectives for computer
literacy in science teachers developed in ENLIST Micros. However,
the Computer Opinion Survey, Version AZ (Maurer and Simonson,
1984) was selected as a good measure of the affective objectives.

Therefore, the co-directors initiated development of a
criterion-referenced test for the cognitive competencies developed
by ENLIST Micros. A criterion-referenced test was selected over
a norm-referenced test because the test is to "yield measurements
that are directly interpretable in terms of specific performance
standards (Glaser and Nitko, 1971, p. 653)."

The purpose of the test is to classify teachers as masters
or non-masters of computer literacy in educational computing in
the sciences. Therefore, the test should be designed to provide
information about the specific knowledge and skills of the examinees
and to facilitate the evaluation of the curriculum designed by
the ENLIST Micros project.

Procedure

A twelve step procedure adapted from the model proposed by
Hambleton (1984, p. 201) for the development and validation of
criterion-referenced tests is being used to develop the Test of

Computer Literacy for Science Teachers (TCLST). The steps are:

1. Preliminary considerations
2. Review of objectives
3. Item writing
4. Assessment of content validity
5. Revisions to test items
6. Field test administration
7. Revisions of test items
8. Test assembly
9. Selection of a standard
10. Pilot test administration
11. Preparation of manuals
12. Additional technical data collection

The first step involved specifying the test purposes, the groups to be measured, and the initial estimate of the test length. The objectives for the test were selected and reviewed as part of a previous study (Ellis and Kuerbis, 1985) to develop and validate the essential competencies in computer literacy in science teachers (step two).

Item writing (step three) followed procedures established by Popham (1980) for test item specification. The domain specifications included a description of the purpose of the instrument, the definition of computer literacy for science teachers, a list of the essential competencies, a description of each item type, and sample items. The writers were encouraged to develop higher-order questions (Bloom, Engelhart, Furst, Hill, and Krathwohl, 1956). The items were edited and revised by the project co-directors.

Berk's (1984) procedures for selecting items for a criterion-referenced test were followed in this study. A panel of 12 experts in educational computing in the sciences examined each item and its respective objective to make a judgment about the item-objective congruence and the technical quality of each item (step four). Items were revised by the co-directors according to

4

the recommendations of the panel of experts and in consideration of possible content bias related to culture, sex, race, or age (step five).

The items were administered to a group of science teachers who were novices in educational computing and to the same group after instruction with the materials developed by ENLIST Micros (step six). In addition, some science teachers provided written and oral reviews of the test items.

Three procedures of statistical analysis were used to analyze the results of the field test administration of the instrument. Item difficulty was calculated for the preinstruction and postinstruction groups as the proportion of teachers who answered the item correctly. Item discrimination was calculated as the proportion of teachers who answered an item correctly on the posttest minus the proportion who answered it correctly on the pretest. For multiple choice items, an item analysis of choice response patterns--the proportion of teachers selecting each distractor--was performed to assist in the revision of the items.

Once the analysis of the field test administration of the instrument was completed, a process of judgmental review of the items was used to select and revise items for the test (step seven). In Table 1, are guidelines proposed by Berk (1984) for selecting items for criterion-referenced tests considered in this study:

Table 1
Guidelines for Selecting Criterion-referenced Test Items

Item Characteristic	Criterion	Index Value
Item-objective congruence	Matches objective being assessed	None
Difficulty	Difficult for uninstructed group	0-50
	Easy for instructed group	50-100
Discrimination	Positively discriminates between criterion groups	Positive .10-1.00

However, as Berk (1984, p. 123) emphasizes, selecting only the best discriminating items "tends to maximize decision validity at the expense of content validity." Items with low discrimination indices but with high objective congruence would be discarded if only statistical data were considered.

> The range of values should be viewed from the perspective of what seems reasonable relative to the content and the importance of the behavioral objectives, the ability and background characteristics of the students, and the actual instruction to which the students were exposed. (Berk, 1984, p. 123)

Since for all items the item-objective congruence had been acceptable, those items meeting the minimum criteria for item discrimination in Table 1 were considered for inclusion in the test without revision.

However, a further analysis was used to examine all multiple choice items considered for use in the instrument. Item analysis of choice response patterns for multiple choice items were compared to the following criteria:

1. Each distractor should be selected by more students in the uninstructed group than in the instructed group.

2. At least a few uninstructed students (5-10%) should choose each distractor.

6

3. No distractor should receive as many responses by the instructed group as the correct answer.

(Berk, 1984, p. 127)

The items from the selection and revision processes were combined to form the TCLST (step eight). The TCLST instrument includes test directions, practice questions, test items, and a scoring key.

Four steps remain in the procedure to develop the TCLST-- selection of a standard, pilot test administration, preparation of manuals, and additional technical data collection. To complete these steps the TCLST will be administered to another two groups of science teachers, experts and novices. Both groups will include science teachers from all grade levels and from all science disciplines. The group of masters will be science teachers who have completed training in educational computing in the sciences and who have used computers to enhance science teaching for at least one year. The group of non-masters will be science teachers who have never completed training in computer literacy and who have never used computers to enhance science teaching.

Using the results of administration of the TCLST to those groups, the co-directors will select a standard to identify masters and non-masters of educational computing in the sciences (step nine) and will determine the reliability and validity of the instrument (step 10). Once the test is determined to be reliable and valid and to be effective for identifying masters of educational computing in the sciences, the co-directors will prepare manuals describing testing and analysis procedures (step 11).

7

The last step in the procedure to develop the instrument is to conduct additional reliability and validity investigations using a variety of populations. The co-directors will release the TCLST to other researchers to conduct those investigations.

Results

Step one: preliminary considerations

The test purposes were determined as part of the ENLIST Micros project. The test is to be used as a measure of mastery of the essential competencies of computer literacy in science teachers established by the project. The co-directors specified that the test should contain no more than 20 questions to minimize the time for administration. It was determined that the test would have one question corresponding to each of the 19 competencies from the cognitive domain. The Computer Opinion Survey (Maurer and Simonson, 1984) was selected to measure the competencies from the affective domain.

The co-directors selected the preinstruction-postinstruction method for gathering data to analyze the items. The preinstruction groups were science teachers participating in 12 workshops in different regions of the country and representing both preservice and inservice situations, both elementary and secondary levels of teaching, and a variety of fields of science teaching. The postinstruction groups were the same teachers after participating in a workshop with the ENLIST Micros materials for a minimum of one day--consisting of reading the text, viewing videotape programs and running computer software developed to illustrate educational computing in the sciences, interacting in

8

small groups, searching for software in catalogs, and evaluating commercial software.

For the item analysis, 119 science teachers met the criteria stated above. Forty-six percent of the participants were preservice teachers, 30 percent were elementary teachers, and 22 percent were secondary teachers. The majority of the secondary teachers taught either general science or life/biology science and the remainder of the secondary teachers were evenly distributed across the other science disciplines. All of the participants included in the analysis indicated they were either non-users or novices in educational computing and none had used the computer in science teaching. However, 66 percent indicated they had used a microcomputer for some other purpose. Twenty eight of the participants had advanced degrees, mostly a Masters degree in education.

Step two: review of objectives

A five step process was employed to establish and validate the essential computer literacy competencies for science teachers. A comprehensive list of 160 competencies, suggested by experts in educational computing, were consolidated into a list of 63 competencies by the co-directors. The list of 63 competencies was prioritized by 46 experts in educational computing, 146 science teachers, and 65 elementary principals. The 63 competencies were reduced to 22 essential competencies using the criterion that for the competency to be retained 75 percent or more of the experts rated it as important.

Using factor analysis, the 22 essential competencies were

9

separated into six scales: Computer Awareness, Applications of Microcomputers in Educational Settings, Implementation of Microcomputers in Curriculum and Instruction, Evaluation of Microcomputer Software, Resources in Educational Computing, and Values and Attitudes Toward Educational Computing. The 19 competencies for the five cognitive scales were used as the objective for the development of the TCLST.

Step three: item writing

Using the item specifications developed for the test, the six writers and the co-directors prepared a pool of 74 items. There were at least two items written for each competency.

Step four: assessment of content validity

For each item, a panel of 12 experts in educational computing evaluated the item-objective congruence and its technical quality. Rating scales like the one used by Torardi (1984) were used to determine how well each item represented the competency it was intended to measure (1=poor, 2=fair, 3=average, 4=good, 5=excellent) and to measure the technical quality of the item by indicating whether the item should be accepted or rejected (3=accept without revisions, 2=accept with revisions, 1=reject). An item was considered for further study if the mean rating for item-objective congruence was greater than 3.0 and if 75 percent of the experts did not recommend its rejection based on technical quality.

Step five: revisions to test items

The results of the assessment of content validity, were used by the co-directors to revise the items. After revision, forty items were retained for the 19 objectives. To shorten the length

of the test, the items for each objective were randomly assigned
to one of two forms of the test, thus creating parallel forms
each containing 20 items.

Step six: field test administration

The test forms were administered to the science teachers
(described in results for step one) prior to and following instruc-
tion with materials developed by the ENLIST Micros project.

In the analysis of the responses to the items, if a partici-
pant left an item blank, it was recorded as an incorrect response.

> Since a criterion-referenced test assesses how much a
> student knows unaffected by speed of response...it can
> be assumed that if a student did not respond to an
> item, it was because he or she did not know the answer.
> (Berk, 1984, p. 104)

Item difficulties ranged on the pretests from .925 to .089
and on the posttests from 1.00 to .275 (see Table 2). Only 22.5
percent of the items on the pretests had an item difficulty at or
below .50. However, 80 percent of the items on the posttests had
an item difficulty above .50. The range of discrimination values
was between -.249 and .332. Of the 40 items, 12 had a negative
discrimination value, 15 had positive values below .10, and 13
had values between .10 and 1.00.

An analysis of response patterns for item choice found
that 57.7 percent of the multiple choice items failed to meet
Berk's (1984) criterion that each distractor should be selected
by more students in the uninstructed group than in the instructed
group. Also, 76 percent failed to meet the criterion that at
least a few uninstructed students (5-10 percent) should choose
each distractor. However, only 7.7 percent failed to meet the

11

Table 2
Item Difficulty and Item Discrimination Indices
for Items for the Essential Competencies in the
Cognitive Domain

Competency	Item	Type	Item Difficulty Pre	Post	Item Discrimination	Decision
1	A	MC	.804	.575	-.229	Delete
	B	MC	.80	.88	.08	Revise
2*	A1	MC	.946	.975	.029	Delete
	A2	MC	.696	.85	.154	Retain
	B1	MC	.925	.676	-.249	Delete
	B2	MC	.825	.794	-.031	Delete
3	A	T/F	.911	.95	.039	Delete
	B	MC	.825	.824	-.001	Rewrite
4	A	T/F	.946	.95	.004	Delete
	B	MC	.925	.912	-.013	Rewrite
5	A	MC	.982	1.00	.018	Delete
	B	MC	.875	1.00	.125	Retain
6	A	MC	.786	.60	-.186	Delete
	B	MC	.40	.382	-.018	Rewrite
7	A	MC	.518	.85	.332	Retain
	B	MC	.55	.706	.156	Delete
8	A	MC	.143	.275	.132	Delete
	B	MC	.125	.294	.169	Retain
9	A	MC	.786	.875	.089	Delete
	B	MC	.525	.618	.093	Revise
10	A	T/F	.679	.85	.171	Retain
	B	T/F	.65	.529	-.121	Delete
11	A	MC	.679	.55	-.139	Delete
	B	MC	.35	.412	.062	Revise
12	A	T/F	.786	.90	.114	Retain
	B	T/F	.825	.853	.028	Delete
13	A	T/F	.875	.95	.075	Delete
	B	T/F	.90	.971	.071	Revise
14	A	MC	.589	.575	-.014	Rewrite
	B	MC	.725	.529	-.196	Delete
15	A	MC	.321	.35	.029	Revise
	B	T/F	.225	.294	.069	Delete
16	A	T/F	.839	.925	.086	Delete
	B	T/F	.85	.971	.121	Retain
17	A	T/F	.821	.90	.079	Delete
	B	MC	.225	.382	.157	Retain
18	A	T/F	.821	.725	-.096	Delete
	B	T/F	.90	.912	.012	Revise
19	A	MC	.089	.30	.211	Retain
	B	MC	.175	.50	.325	Delete

*Four items were tested for objective two.

MC=Multiple Choice
T/F=True/False

criterion that no distractor should receive as many responses by the instructed group as the correct answer.

Step seven: revisions to test items

A judgmental review of items was performed using the results of the analyses of item-objective congruence, item difficulty, item discrimination, and response patterns of item choice. For each objective, the item was selected for inclusion in the TCLST that best met the criteria, except where no item met the minimum standard of a positive discrimination value.

In Table 2, the decisions as to which item should be included in the TCLST for each competency are listed. Nine items met the criteria and were selected for retention with minor revision. Six items having low positive discrimination values were selected but required more thorough revision. For four objectives, no items were selected and therefore, new items were written for these objectives. The response patterns for item choice were examined for multiple choice items selected for the TCLST to determine which distractors for an item needed revision.

Step eight: test assembly

The 19 items resulting from the revision procedure were combined to form the TCLST. Included in the instrument were test directions, items to gather descriptive information on the respondent, instructions for answering the items, and 19 test items. An answer key for the test was developed for test administrators.

Discussion

After field testing the curriculum developed in ENLIST Micros and field testing the items for the TCLST, the co-directors deemed the 19 essential competencies of computer literacy in science teachers appropriate. However, some competencies reflect behaviors that are becoming more common among all science teachers, such as those found in the area of computer awareness. In our study, 66 percent of the teachers, who had indicated no use of the computer in science teaching, also indicated they had used a microcomputer for some other purpose. Knowing how to operate a computer still will be an essential competency for defining computer literacy for science teachers, even though, most teachers will have achieved that skill soon.

Because many science teachers already know how to use a microcomputer, many have achieved a beginning level of computer literacy. However, computer literacy for a science teacher is different than general computer literacy. As Bork (1985, p. 34) stated, "the teacher...needs to understand fully the educational role of the computer; the pupil does not." The competencies needed to use the computer to enhance teaching and learning in science are unique and are not learned as part of the general skills often used to define computer literacy.

Difficulties arise in defining computer literacy for science teachers because there is such a diversity of levels of computer awareness among this group. For many items developed in this study, the correct answer was given by more than 85 percent of the preinstruction groups of science teachers This

14

high level of knowledge among the preinstruction groups indicates that those science teachers had prior experience, that some competencies may be too fundamental, or that those items were too easy.

The co-directors decided that, where appropriate, those items having low item difficulties on the pretest should be revised to assess higher levels of knowledge. The findings that many teachers are developing computer awareness and therefore that some of the essential competencies are being achieved by those teachers without formal instruction do not lesson the importance of those competencies. Rather, those essential competencies related to computer awareness should be considered minimum competencies and our goal is to achieve 100% mastery of those competencies by all science teachers.

According to Berk (1984), the analysis of a minimum competency test should use modified guidelines. Moreover, Berk suggests that item discrimination may not be a meaningful statistic for this type of test.

> It is implicit in the measurement of minimum skills.. that the students in the criterion success group perform 100 percent on each item. Students in the second criterion group (incompetents) chosen on the basis of their inability to demonstrate success on the object- ives should perform less than 100 percent on each item (item difficulty < 100 percent). These item difficulty values might be expected to vary between zero and 94 percent. Given that possible range, the magnitude of the discrim- ination indices could also be expected to exhibit wide variability. (Berk,1984, p. 124)

Therefore, if a determination has been made that the competency is essential in defining the domain and that it is a minimum

15

competency, it can be expected that items used to measure that competency may not discriminate well between masters and non-masters.

Other essential competencies in computer literacy in science teachers are not related to computer awareness and are not learned as part of a general introduction to using the computer. These competencies are in the areas of application, implementation, evaluation, and resources of educational computing in the sciences. Even though there is a component to applications and resources common to computer literacy in general and to computer literacy in science teachers, some knowledge and skills are unique to science education. Also, an overlap exists between knowledge and skills in implementation, evaluation, and resources common to general education and those appropriate to educational computing in the sciences. However, it is expected that higher discrimination indices will be found for items measuring knowledge and skills in those areas that are unique to educational computing in the sciences and therefore, that the TCLST will prove useful in classifying teachers as masters or non-masters of this domain and for evaluating programs to train science teachers to use the computer to enhance the teaching and learning of science.

Implications

This study points out the importance of using procedures for developing criterion-referenced tests in many investigations in science education. Too often research studies and evaluations in science education select previously developed norm-referenced tests to determine the difference between groups. Even though a

16

norm-referenced test is designed to maximize reliability and
therefore increases the likelihood of finding statistically
significant differences between treatment groups, it may not
explain the level of mastery by the post-treatment group of
the desired competencies.

This study, illustrates the difficulties of developing a
criterion-referenced test, including specifying valid objectives,
establishing good item-objective congruence, and especially
writing items that discriminate well between masters and non-mas-
ters. However, the procedures used in this study for developing
criterion-referenced tests are very effective at identifying the
weaknesses in the items and these procedures will ultimately lead
to an effective instrument.

When the development process is completed, the TCLST will be
a useful tool in science education for measuring the mastery of
computer literacy in science teachers. A cutoff point for mastery
will be established and can be used for comparisons of treatment
groups. Perhaps a good method for analyzing the effect of the
treatment will be to compare the proportion of subjects classified
as masters in the preinstruction group with the proportion of
subjects classified as masters in the postinstruction group.
After comparing treatment groups, a more detailed analysis of
mastery of individual objectives can point out areas where the
curriculum can be improved and where remediation is needed.

17

References

Berk, R.A.   A guide to criterion-referenced test construction.
    Baltimore: The Johns Hopkins University Press, 1984.

Bloom, B.S., Engelhart, M.D., Furst, E.J., and Hill, W. H., and
    Krathwohl, D.R.   Taxonomy of educational objectives, the
    classification of educational goals--Handbook I: cognitive
    domain.   New York: David McKay Company, Inc., 1956.

Bork, A.   Personal computers for education.   New York: Harper
    and Row Publishers, 1985.

Ellis, J.D. and Kuerbis, P.J.   Development and validation of
    essential computer literacies for science teachers.   A paper
    presented at the 1985 annual NARST meeting, French Lick
    Springs, Indiana, April 1985.   ED255 373.

Glaser, R. and Nitko, A.J.   Measurement in learning and instruc-
    tion.   In R.L. Thorndike (Ed.) Educational Measurement   (2nd
    Ed.).   Washington, D.C.:   American Council on Education,
    1971.   pp. 625-670.

Hambleton, R.K.   Validating the test scores.   In Berk, R.A. (Ed.)
    A guide to criterion-referenced test construction.
    Baltimore: The Johns Hopkins University Press, 1984.

Maurer, M. and Simonson, M.R.   Computer Opinion Survey (Version
    AZ).   Ames IA: Iowa State University Research Foundation,
    Inc., 1984.

Popham, W.J.   Domain specification strategies.   In Berk, R.A.
    (Ed.)   A guide to criterion-referenced test construction.
    Baltimore: The Johns Hopkins University Press, 1980.

Taylor, J.L. (Ed.)   Teacher shortage in science and mathematics:
    myths, realities, and research.   Washington, D.C.:   National
    Institute of Education, 1984.

Torardi, M.M.   The development of a computer literacy assessment
    instrument (project report).   College of Education, Iowa
    State University, Ames, IA.

Watson, R.F.   Computer-based projects in science education: An
    overview.   Journal of College Science Teachers, 1983, 12/6,
    373-380.